

AD \_\_\_\_\_

GRANT NUMBER DAMD17-94-J-4371

TITLE: Computer Aided Breast Cancer Diagnosis

PRINCIPAL INVESTIGATOR: Carey E. Floyd, Ph.D.

CONTRACTING ORGANIZATION: Duke University  
Durham, North Carolina 27710

REPORT DATE: October 1998

TYPE OF REPORT: Annual

PREPARED FOR: Commander  
U.S. Army Medical Research and Materiel Command  
Fort Detrick, Frederick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;  
distribution unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

DTIC QUALITY INSPECTED 4

20000111 027

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

|   |   |  |  |  |
|---|---|--|--|--|
| 1. AGENCY USE ONLY (Leave blank)  |   | 2. REPORT DATE<br>October 1998                             | 3. REPORT TYPE AND DATES COVERED<br>Annual (23 Sep 97 - 22 Sep 98) |  |
| 4. TITLE AND SUBTITLE<br>Computer Aided Breast Cancer Diagnosis   |   |  | 5. FUNDING NUMBERS<br>DAMD17-94-J-4371                             |  |
| 6. AUTHOR(S)<br>Carey E. Floyd, Ph.D.   |   |  |  |  |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>Duke University<br>Durham, North Carolina 27710   |   |  | 8. PERFORMING ORGANIZATION<br>REPORT NUMBER                        |  |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)<br>Commander<br>U.S. Army Medical Research and Materiel Command<br>Fort Detrick, Frederick, MD 21702-5012 |   |  | 10. SPONSORING/MONITORING<br>AGENCY REPORT NUMBER                  |  |
| 11. SUPPLEMENTARY NOTES   |   |  |  |  |
| 12a. DISTRIBUTION / AVAILABILITY STATEMENT<br>Approved for public release; distribution unlimited   |   |  | 12b. DISTRIBUTION CODE   |  |
| 13. ABSTRACT (Maximum 200)  |   |  |  |  |
| 14. SUBJECT TERMS Mammography, Radiology, Computer-Aided<br>Diagnosis, Neural Networks, Malignancy Prediction, Artificial<br>Intelligence, Breast Cancer          |   |  | 15. NUMBER OF PAGES<br>14  |  |
|   |   |  | 16. PRICE CODE   |  |
| 17. SECURITY CLASSIFICATION<br>OF REPORT<br>Unclassified  | 18. SECURITY CLASSIFICATION<br>OF THIS PAGE<br>Unclassified | 19. SECURITY CLASSIFICATION<br>OF ABSTRACT<br>Unclassified | 20. LIMITATION OF ABSTRACT<br>Unlimited                            |  |

## FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

Cef Where copyrighted material is quoted, permission has been obtained to use such material.

Cef Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

\_\_\_\_ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

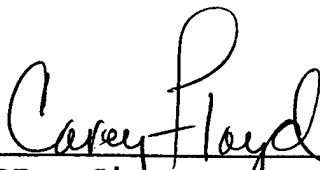
\_\_\_\_ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

\_\_\_\_ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

\_\_\_\_ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

\_\_\_\_ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

\_\_\_\_ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

  
\_\_\_\_\_  
PI - Signature

\_\_\_\_\_  
Date

Computer Aided Breast Cancer Diagnosis  
PI: Carey E. Floyd Jr.

|  | Contents |    |
|--|----------|----|
| Report Documentation Page                    |          | 1  |
| Foreword                                     |          | 2  |
| Table of Contents                            |          | 4  |
| Summary of progress on the Statement of Work |          | 5  |
| Body of Report                               |          | 7  |
| References                                   |          | 14 |

The long range goal of this project is to improve the accuracy and consistency of breast cancer diagnosis by developing a Computer Aided Diagnosis (CAD) system for early prediction of breast cancer from patients' mammographic findings and medical history. Specifically, this system will predict the malignancy of non-palpable lesions that are examined with diagnostic mammography and are considered for biopsy. The goal is to improve the specificity of diagnosis with little loss of sensitivity thus significantly improving the positive predictive value of breast biopsy.

Toward this goal, we have developed an artificial neural network (ANN) to predict biopsy outcome from mammographic and history findings. In the first four years of the grant we have 1) developed a user interface for acquiring mammographic findings, 2) acquired 700 cases using the standardized BI-RADS<sup>TM</sup> reporting system, 3) trained and evaluated several ANN predictive models, 4) conducted a small prospective study, 5) examined the inter- and intra-observer variability of the reporting lexicon, 6) investigated reducing the number of active input features, and 7) examined the sensitivity of the system to the techniques used for sampling the data.

What follows is a point by point assessment of the progress for each task in the original statement of work:

#### **Statement of Work**

Task 1, Develop an ANN to predict biopsy outcome from mammographic and history findings.

Years 1-4

Development will start with the successful preliminary backpropagation network.

The significant improvements needed include: 1) larger set of clinical cases to better represent the general patient population, 2) higher specificity while maintaining >98% sensitivity. The preliminary work will be extended as follows.

Year 1

1.1) Expand the number of input features, both mammographic and medical history.

The ANN will be implemented on a workstation (SUN SPARC) to allow the size

of the network to be enlarged. This will allow more medical history and radiological features to be included.

**These tasks were all achieved in year one.**

#### Year 2-4

1.2) Develop a time-series ANN to examine current as well as previous exams.

**Note: this aim was dropped in response to the decreased budget as negotiated with BC Baker in a revised statement of work in August 1994 .**

1.3) Evaluate other ANN architectures which have been demonstrated to be appropriate for pattern classification.

**Achieved in year 2.**

#### Year 3-4

2) Evaluate the improvement in radiologists' diagnostic performance when the computer diagnostic aid is provided.

#### Year 3

2.1) Install the trained network on the Mammography Database server to perform on-line prediction as the radiologists input the features.

**Achieved in year 2.**

#### Year 3-4

2.2) Test the hypothesis that use of the network prediction by radiologists will increase diagnostic accuracy (prediction of biopsy results).

**Not yet achieved. Begun in year 3.**

In summary, all of aim 1 has been achieved and previously reported. All that remains is to finish the evaluation of the improvement in radiologists' performance when the system is used. This work was slightly delayed by a change in directorship of the division of mammography and by a change of management in the radiology informatics group. The only real difficulty was a result of the change in the informatics

system. A workaround has been initiated and the evaluation is scheduled for the spring of 1999.

This report describes a computer aid to predict the malignancy of non-palpable lesions that are examined with diagnostic mammography and are considered for biopsy. The goal is to improve the specificity of diagnosis with little loss of sensitivity thus significantly improving the positive predictive value of breast biopsy. An artificial neural network (ANN) is described to assist radiologists in the differentiation of benign from malignant lesions. Inputs to the ANN were derived from the patient's history and the radiologist's description of lesion morphology following the ACR Breast Imaging Reporting and Data System (BI-RADS<sup>TM</sup>). The output of the neural network is the likelihood of malignancy. Evaluation of the system on 500 cases demonstrates that 22% of the benign biopsies could be avoided without missing a malignancy. At this threshold, the positive predictive value (PPV) of biopsy would be improved from 35% to 41%. With a less conservative approach, 41% of the benign biopsies could be avoided while still performing biopsies on 98% of the malignancies. At this threshold, the positive predictive value (PPV) of biopsy would be improved from 35% to 47%.

## 1. INTRODUCTION

The lifetime risk of developing breast cancer has increased to one woman in eight <sup>1</sup>. While screening mammography can decrease the mortality due to breast cancer by 30%<sup>2,3</sup>, improvements in the diagnosis are still needed. Although mammography is a sensitive tool for detecting breast cancer, the positive predictive value (PPV) is low <sup>4-6</sup>. Several factors contribute to this, including similarity in the radiographic appearance of benign and malignant breast lesions <sup>6</sup> as well as an overall conservative approach of physicians <sup>7</sup>. Only 10-34% of women who have biopsy for mammographically suspicious nonpalpable lesions have a malignancy by histologic diagnosis<sup>5</sup>. Currently, more than a million biopsies are performed each year <sup>8</sup>. Due to the present low PPV of mammography, hundreds of thousands of women undergoing biopsy for a benign finding are unnecessarily subjected to the discomfort, expense, potential complications, change in cosmetic appearance, and anxiety that can accompany breast biopsy <sup>5,9-11</sup>.

The long range goal of this project is to improve the accuracy and consistency of breast cancer diagnosis by developing a Computer Aided Diagnosis (CAD) system for early prediction of breast cancer from patients' mammographic findings and medical history. Specifically, this system will predict the malignancy of non-palpable lesions that are examined with diagnostic mammography and are considered for biopsy. The

goal is to improve the specificity of diagnosis with little loss of sensitivity thus significantly improving the positive predictive value of breast biopsy.

Toward this goal, here we describe the development of an artificial neural network (ANN) to assist radiologists in the differentiation of benign from malignant lesions. Inputs to the ANN were derived from the patient's history and the radiologist's description of lesion morphology following the ACR Breast Imaging Reporting and Data System (BI-RADS<sup>TM</sup>). The output of the neural network is the likelihood of malignancy.

Artificial neural networks are a form of artificial intelligence analogous to layers of biological neurons. These networks can be trained to "learn" essential information from a set of data. The structure of an ANN is a set of processing units (nodes) arranged in rows. Input nodes are interconnected by simple calculations with an internal layer of hidden nodes and a single output node. Rather than having a fixed algorithmic approach to a classification problem, an ANN is sequentially presented with a set of supervised training cases — input data paired with the correct output. The ANN modifies its behavior ("trains") by adjusting the strength or "weights" of the connections until its own output converges to the known correct output. The information "learned" by the ANN is stored in the weight the network gives to connections between nodes.

## 2. METHODS

The ANN for prediction of breast malignancy was constructed as a three layer feed-forward network with a backpropagation training algorithm. The layers consist of an input layer with 18 input nodes, one hidden layer with 10 nodes, and an output layer with one output node. Each input node corresponds to either a radiologist's description of a feature of the lesion or information from the patient's medical or family history.

A total of 500 lesions were identified on mammograms of those women undergoing needle localization for nonpalpable breast lesions that went on to open excisional biopsy and pathological diagnosis. Each mammographic study was acquired using film-screen technique on dedicated mammography equipment. No case was included in the study if either of the reviewing radiologists had prior knowledge of the biopsy results or if the suspicious area was not definitely identified. Of the 500 lesions evaluated, there were 232 masses alone, 192 suspicious calcifications, and 29 combinations of masses and associated microcalcifications. The remaining 47 lesions included architectural distortion, regions of asymmetric breast density, areas of focal asymmetric density, and areas of asymmetric breast tissue. Patients ranged in age from 24 to 86 years with an average age of 55 years. At biopsy, 326 (65%) of the lesions were found to be benign while 174 (35%) were malignant. This PPV of 35% is somewhat greater than that described in prior studies.

Each set of training films was reviewed prospectively by one of two radiologists whose primary clinical responsibilities are the interpretation of mammograms and the evaluation of breast lesions and who are familiar with the definitions of the BI-RADS<sup>TM</sup> descriptors. At least two views of the breast with the suspicious lesion were provided to the participating radiologists; a cranio-caudal and mediolateral-oblique view were available in all cases. Other views including true lateral, magnification



views, and spot compression views as well as comparisons with the opposite breast were provided for evaluation when available. In order to avoid biasing the radiologist's description of the lesion, films from prior studies and the patient's history were initially withheld while the reviewing radiologist chose descriptors for each lesion. The radiologist described each lesion using the BI-RADS<sup>TM</sup> lexicon by completing a checklist that included all possible BI-RADS<sup>TM</sup> descriptors. The reviewing radiologist selected only a single descriptor from each category. Each reader was blinded to the biopsy results while reviewing the films.

There were 18 inputs to the ANN. Ten of the inputs were morphologic features assigned to the mammographic image of the lesion by a radiologist. Eight of the inputs were from the patient's personal and family history. These data were from a survey form completed by the patient at the time of the exam. Each input is information routinely collected using the ACR BI-RADS<sup>TM</sup> standardized lexicon.

Three of the features, calcification distribution, number and description, apply to microcalcifications and calcifications associated with masses. Four of the features apply only to masses: mass margin, shape, density, and size.

The patient's history provides the other 8 inputs. These include the patient's age, history of prior breast cancer, history of prior ipsilateral benign biopsy, family history of breast cancer, menstrual status, and use of estrogen or progesterone therapy. All mammographic features and patient history findings were assigned a numerical value scaled so that each input ranged from zero to one. The scaling of the inputs was selected after discussion with experienced mammographers and a review of the literature concerning the BI-RADS<sup>TM</sup> descriptors.

### 3. RESULTS

The classification performance of the model is shown below in Figs. 1 and 2 as histograms of the benign and malignant cases binned by the ANN model output. If a threshold is set between two bins, the cases to the left of the threshold will be called benign while the cases to the right will be called malignant. The shaded bars represent the benign cases and the solid bars represent the malignant cases. In fig. 1, the model shows good behavior for the benign cases as seen by the predominant grouping to the left. Performance for the malignant cases is not as dramatic but still results in a good separation of the two classes. It is evident that setting a threshold at around 0.1 will save over 100 benign biopsies while missing few malignancies. To examine this region further, the histogram is expanded in fig. 2 to show the region between model outputs of 0 and 0.1. In this region now we see that a threshold can be set to save some benign biopsies while missing no malignancies. The performance of the network as the decision threshold is varied is shown in Table 1.

It is common to report the performance of classification models using Receiver Operating Characteristic plots as shown in Fig. 3. In addition, it is common to show fitted ROC curves based on a normal model of the histograms. From examination of the histogram in Figs. 1 and 2, it is evident that while the malignant cases could be represented by a normal distribution, the benign cases could not. Indeed, when fitted using a normal model, the left hand region of the histograms is poorly fit. This is unfortunate since this is the high sensitivity region that is of most interest for cancer diagnosis models. For this reason, we show the ROC curve computed from the data

case by case in Fig. 3. The sensitivity, Specificity, and positive predictive value are shown for one threshold. The area  $A_z$  is computed from Newton's method.

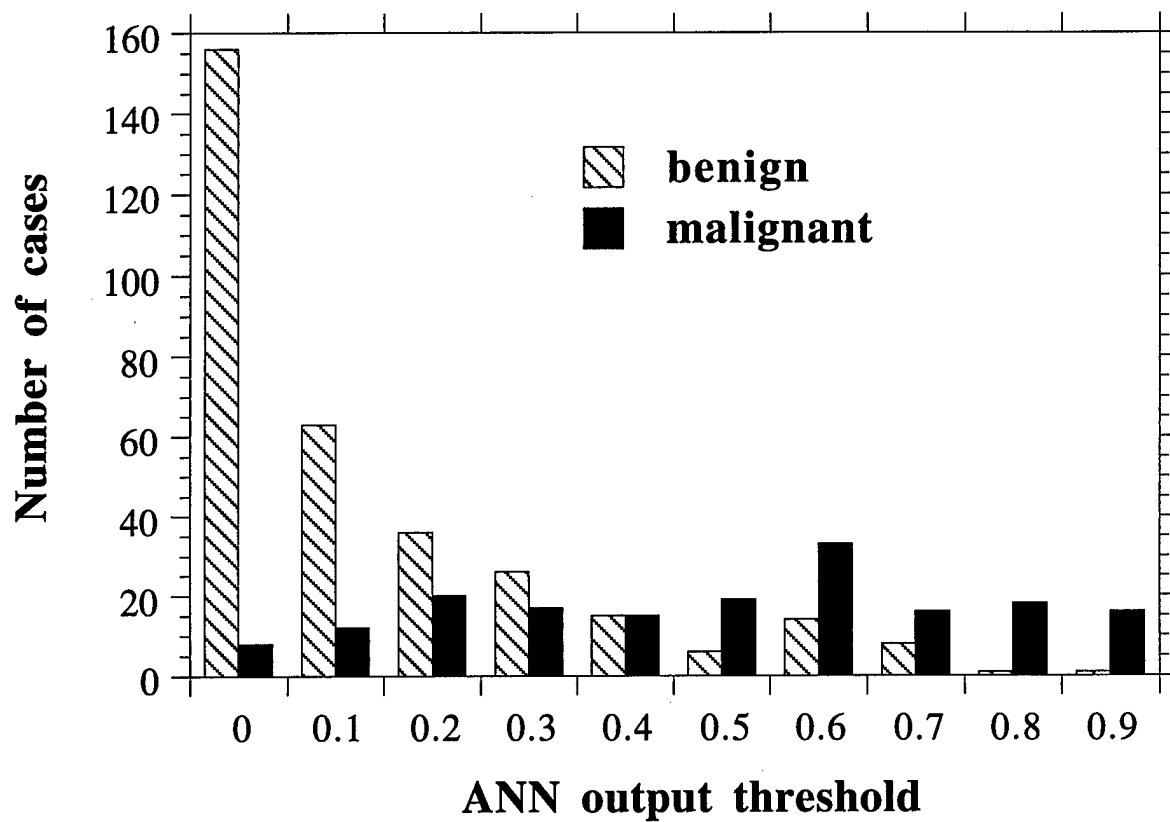


Fig. 1. Histogram of cases binned by ANN model output.

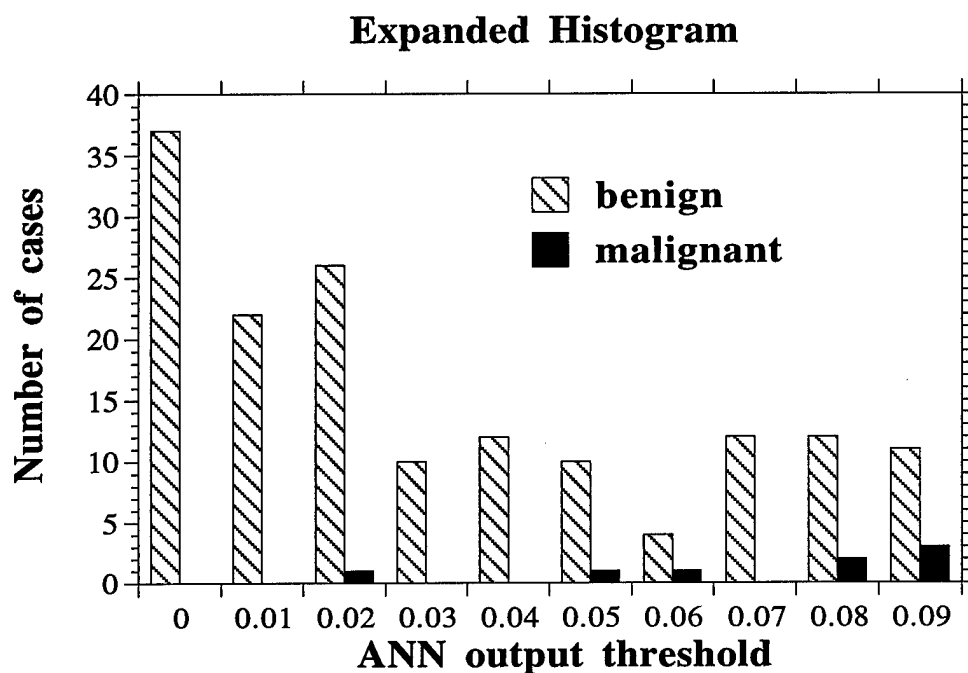


Fig. 2. Histogram expanded to emphasize the region between 0 and 0.1.

Table 1

Performance of the trained neural network

| Sensitivity | Specificity | Performance: Sparing Benign Biopsies |                     |                        | ANN Output Threshold |
|-------------|-------------|--------------------------------------|---------------------|------------------------|----------------------|
|             |             | Positive Predictive Value            | Malignancies Missed | Benign Biopsies Spared |                      |
| 100         | 0           | 35                                   | 0                   | 0                      | 0.000                |
| 100         | 22          | 41                                   | 0                   | 72                     | 0.025                |
| 98          | 41          | 47                                   | 4                   | 133                    | 0.081                |
| 95          | 52          | 51                                   | 9                   | 168                    | 0.119                |
| 90          | 64          | 57                                   | 17                  | 208                    | 0.175                |
| 85          | 69          | 59                                   | 26                  | 225                    | 0.216                |

For the decision to biopsy, another important way to visualize the model performance is to plot the number of benign biopsies that would be saved or avoided along with the number of malignancies that would be missed as a function of the decision threshold. These are plotted in Fig. 4. The solid line represents the number of benign biopsies that would be saved while the dashed curve represents the number of malignancies that would be missed as the threshold is varied.

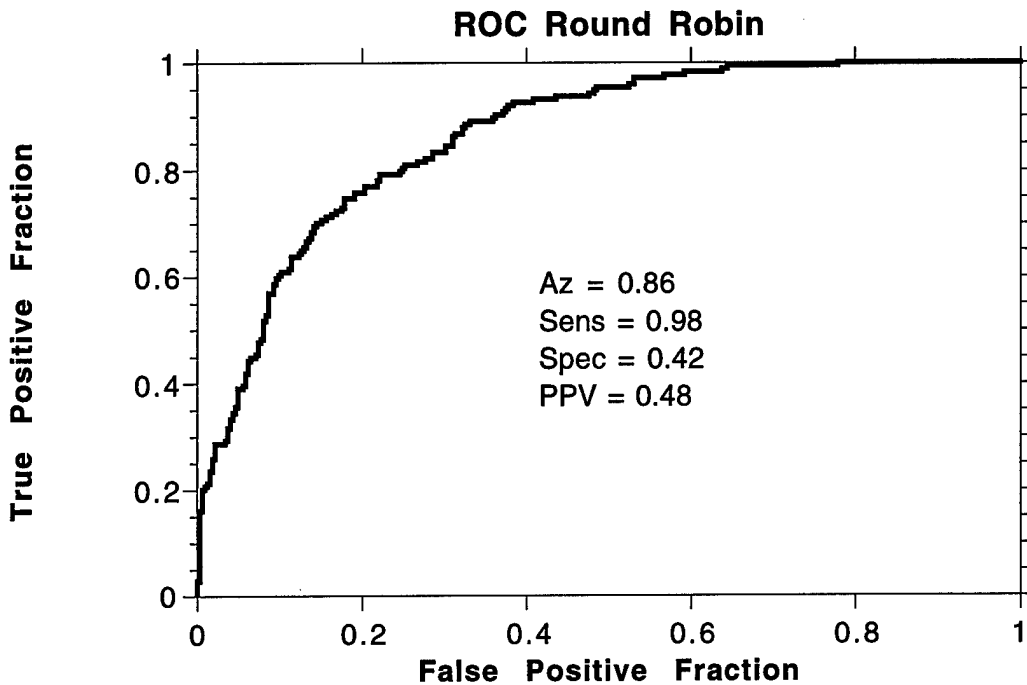


Fig. 3. Receiver Operating Characteristic plotted from the data case by case.

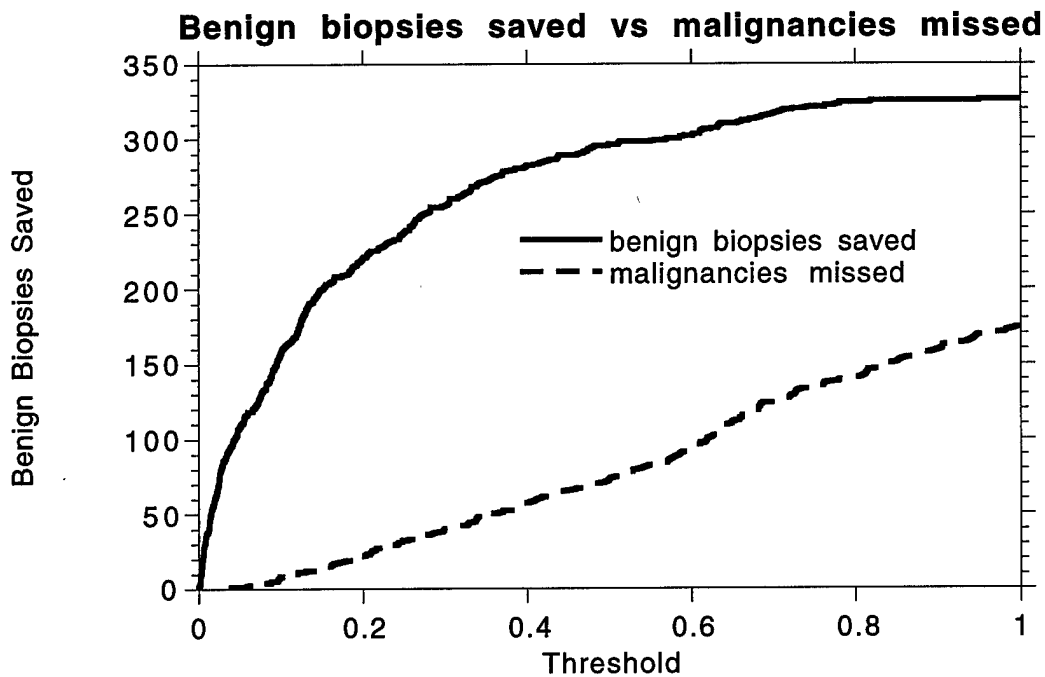


Fig. 4. The number of benign biopsies that would be saved or avoided along with the number of malignancies that would be missed as a function of the decision threshold

## 4. DISCUSSION

Previous work includes rule-based systems<sup>12</sup>, neural network approaches by others<sup>13</sup>, and ourselves<sup>14-17</sup>, and recent work using Bayesian networks<sup>15</sup>. One of the most important aspects of the problem has not been addressed in this work. This is the relative cost of saving a benign biopsy compared with the cost of missing a malignancy. This is critical to selecting an operating point for the decision threshold. If the costs are equal, then a threshold of 0.8 will be optimal. The costs are not equal and clearly the cost of missing a malignancy is greater than the cost of performing a benign biopsy. If the cost of missing a malignancy is infinite and the cost of a biopsy is zero, then the decision level should be set at 0.25 (from table 1) and the system would still save 22% of the benign biopsies. The cost analysis must include "quality of life" measures which are difficult to estimate and measure. Further, the cost is dependent on the proposed treatment strategies. If all cases that are called benign by this system are followed closely, then the cost of missing a malignancy at this stage will be less than if the patient is simply returned to the screening pool. Cost analysis for this project is underway. Other future work will include clinical trials and evaluation of the place for such a computer decision aid in the diagnosis and treatment plan for breast cancer patients.

## 5. REFERENCES

1. Garfinkel, L, Boring, C C, and Heath, C W. Changing trends: an overview of breast cancer incidence and mortality. *Cancer* 1994; 74: 222-227.
2. Shapiro, S. Screening: assessment of current studies. *Cancer* 1994; 74: 231-238.
3. Verbeek, A L M, Hendriks, J H C L, Holland, R, Mravunac, M, Sturmans, F, and Day, N E. Reduction of breast cancer mortality through mass screening with modern mammography. *Lancet* 1984; 1: 1222-1224.
4. Adler, D D and Helvie, M A. Mammographic biopsy recommendations. *Current Opinion in Radiology* 1992; 4: 123-129.
5. Kopans, D B. The positive predictive value of mammography. *American Journal of Roentgenology* 1992; 158: 521-526.
6. Ciatto, S, Cataliotti, L, and Distanti, V. Nonpalpable lesions detected with mammography: review of 512 consecutive cases. *Radiology* 1987; 165: 99-102.
7. Hall, F M. Screening mammography - potential problems on the horizon. *New England Journal of Medicine* 1986; 314: 53-55.
8. Hall, F M, Storella, J M, Silverstone, D Z, and Wyshak, G. Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography. *Radiology* 1988; 167: 353-358.
9. Schwartz, G F, Carter, D L, Conant, E F, Gannon, F H, Finkel, G C, and Feig, S A. Mammographically detected breast cancer: nonpalpable is not a synonym for inconsequential. *Cancer* 1994; 73: 1660-1665.
10. Helvie, M A, Ikeda, D M, and Adler, D D. Localization and needle aspiration of breast lesions: complications in 370 cases. *AJR* 1991; 157: 711-714.
11. Dixon, J M and John, T G. Morbidity after breast biopsy for benign disease in a screened population. *Lancet* 1992; 1: 128.
12. Getty, D. J., R. M. Pickett, et al. Enhanced interpretation of diagnostic images. *Investigative Radiology* 1988; 23: 240-252.
13. Wu Y, Giger M, Doi K, Vyborny C, Schmidt R, and Metz C. Artificial neural networks in mammography: application to decision making in the diagnosis of breast cancer. *Radiology* 1993; 187: 81-87.
14. Floyd, C. E., Jr, J. Y. Lo, et al Prediction of breast cancer malignancy using an artificial neural network." *Cancer* 1994; 74(11): 2944-2948.
15. Baker, J A, Kornguth, P J, Lo, J Y, and Floyd, C E. Artificial Neural Network: Improving the quality of breast biopsy recommendations. *Radiology* 1996; 198: 131-135.
16. Baker, J A, Kornguth, P J, Lo, J Y, Williford, M E, and Floyd, C E, Jr. Artificial neural network for the prediction of breast cancer using BI-RADS™ standardized lexicon. *Radiology* 1995; 196: 817-822.
17. Lo, J Y, Baker, J A, Kornguth, P J, and Floyd, C E, Jr. Computer-aided diagnosis of breast cancer: artificial neural network approach for optimized merging of mammographic features. *Academic Radiology* 1996; 2: 841-850.
18. Kahn, C. E., Jr., L. M. Roberts, et al. Construction of a Bayesian network for mammographic diagnosis of breast cancer. *Computers in Biology & Medicine* 1997; 27(1): 19-29.